The Genetics of Developmental Disorders

Elizabeth J. Radford [1,2,3], Helen V. Firth[3,4]

Job titles:

EJR: Academic Clinical Lecturer in Paediatric Neurology, University of Cambridge.

HVF: Consultant Clinical Geneticist at Cambridge University Hospitals; Honorary Faculty Member of the Wellcome Sanger Institute; Bye-Fellow of Newnham College, University of Cambridge

Academic Degrees:

EJR: BA MB BChir PhD MRCPCH

HVF: DM, FRCP, DCH

Correspondence:

EJR: ejr41@cam.ac.uk

HVF: hvf21@cam.ac.uk

Corresponding author:

hvf21@cam.ac.uk

Affiliations: [1]Department of Paediatrics, University of Cambridge, Cambridge, UK; [2]Department of Paediatric Neurology, Cambridge University Hospitals NHS Foundation Trust, Cambridge UK; [3]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK; [4] East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge UK

Conflict of interest statement

Neither EJR nor HVF have any conflicts of interest.

Key words

Developmental disorder, genetic, genomic, next-generation sequencing, exome, whole-genome sequencing, data security, genetic testing, consent, ethics.

Abstract

Children with developmental disorders often have complex health needs and require frequent input from many different paediatric services. Genomic technologies have revolutionised how these disorders are diagnosed, and have the real potential to transform the management of previously intractable paediatric conditions. For paediatricians to fully harness these benefits for our patients, and to work productively with our clinical genetics colleagues within the multi-disciplinary team, we must be genomically literate and aware of the technical and ethical challenges. In this review article, we will summarise the current understanding of the genetic architecture of developmental disorders; discuss the different types of genetic tests currently available, their strengths and limitations in clinical practice; and discuss the challenges and future opportunities in paediatric genomic medicine.

Introduction

Developmental disorders are defined as diseases that arise in embryonic life or early fetal brain development and alter the developmental trajectory. In the UK, the majority of developmental disorders are believed to have a genetic aetiology. However, an important proportion are caused by exposure to environmental teratogens such as alcohol, sodium valproate; or *in utero* infection by pathogens such as Toxoplasma, cytomegalovirus or Zika virus. While individual developmental disorders are often extremely rare, collectively they are common: approximately 2-3% of children are born with major congenital anomalies and/or demonstrate severe neurodevelopmental disorders. These children often have complex health needs, and will frequently require paediatric input. Indeed, McCandless and others have found that a third of paediatric inpatient admissions are of children with a genetic underlying disorder, with up to half of acute inpatient healthcare costs attributable to their care. Until recently, many of these children did not receive a specific genetic diagnosis. However, the advent of “next-generation sequencing” (NGS) technologies has driven down the cost of genetic testing and made it available within routine clinical practice. Patient diagnosis and care is best delivered by a multi-disciplinary team, involving clinical

geneticists and paediatricians together with other health professionals. While the role and expertise of the clinical geneticist is more crucial than ever, families correctly expect their paediatricians to be up to date with this fast-developing field. Furthermore, as diagnosis rates increase, we have an opportunity to pursue targeted treatments for individual developmental disorders. Such personalised medicine has the real potential to transform the management of previously intractable paediatric conditions. However, for paediatricians to fully harness these benefits for our patients, we must be genomically literate and aware of the technical and ethical challenges. In this review article, we will discuss the benefits to the child and their family of establishing a genetic diagnosis; summarise the current understanding of the genetic architecture of developmental disorders; discuss the different types of genetic tests currently clinically available, their strengths and limitations; and discuss the challenges and future opportunities in paediatric genomic medicine.

1. Why establish a genetic diagnosis?

There are advantages and disadvantages to pursuing a genetic diagnosis, and the balance between the two will be specific to each family. Different families may therefore make very different decision with regards to whether to pursue genetic diagnosis. Paediatricians are increasingly taking consent for some forms of genetic testing, such as karyotyping and microarray analyses, and we have a responsibility to ensure that this consent is fully informed and therefore valid. The advantages of establishing a specific genetic diagnosis include informing clinical management through condition specific medical and prognostic information; the emotional closure that may come with establishing a diagnosis; accuracy of advice regarding risk to other family members; and access to patient support groups. Indeed, there is evidence that although most developmental disorders do not have curative treatments, obtaining a genetic diagnosis alters clinical management in 70% of cases, by informing symptomatic management and clinical monitoring. Having a specific genetic diagnosis may allow a family to access reproductive counselling if they wish, including, in some cases, pre-implantation or pre-natal genetic diagnosis in future pregnancies. Families often report that access to support groups and a community of similarly affected individuals can be of significant practical help, and can allay the sense of isolation that many families experience when they have a child affected by a rare disorder. However, families must also be warned of the possibility of identifying variants of uncertain significance (VUS) and

incidental findings (see Boxes 1, 2 and Table 2), and it is important to adequately explore with them how they would feel about receiving prognostic information regarding their child's condition. For example, some families would not wish to know that their neonate has a significant risk of developing intellectual disability. For others, there may be a stigma associated with genetic diagnoses. These views must be respected although where a genetic diagnosis has the potential to substantially alter clinical management, this should be carefully explained to ensure that the preference is well-informed and in the child's best interest. See Box1 for more practical details regarding taking consent for genetic testing.

2. The genetic architecture of developmental disorders

The genetic abnormalities that can result in developmental disorders range in scale from chromosomal disorders such as aneuploidies – having missing or extra chromosomes – through to the change of single base, a single nucleotide variant (SNV). There is a corresponding range of genetic tests available, each with a different resolution, and with clinical advantages and disadvantages (Table1). As the resolution of the genetic test increases, so does the volume of data which it generates. This raises analysis, data-management and ethical challenges, including an increasing risk of identifying VUS and incidental findings. High-resolution tests (such as whole-genome sequencing) are also substantially more expensive. Therefore, it is important to choose the most appropriate genetic test for the likely genetic abnormality in a particular clinical presentation, and to counsel parents regarding which genetic anomalies may be confidently excluded by the test, and which cannot.

2.1 Structural variants and uniparental disomy:

2.1.1 Aneuploidies

In aneuploidies there are an abnormal number of chromosomes in each cell. For example, in trisomies there are three copies of a chromosome, in monosomies only one. The commonest aneuploidy is Trisomy 21 (Down's syndrome), with a prevalence in the UK of approximately 1 per 1000 births. Other aneuploidies include Patau's syndrome (Trisomy 13) and Edward's syndrome (Trisomy 18), Klinefelter's syndrome (47,XXY), XYY syndrome (47,XYY) Triple X syndrome (47,XXX) and Turner's syndrome (45,X). Most aneuploidies are caused by non-disjunction – the failure of correct separation of the homologous

chromosomes or sister chromatids during cell division. Non-disjunction events in the early embryo can result in mosaic aneuploidies, where there is a mixture of cells in the embryo, some with the correct number of chromosomes, and others bearing an aneuploidy. Occasionally aneuploidies are caused by translocation events – where part of one chromosome becomes attached to another. This is important because an unaffected parent may carry a Robertsonian or a balanced translocation – where they have the correct amount of genetic information, but it is distributed in an unusual way. For example, translocation of the long arm of chromosome 21 to the long arm of chromosome 14 results in a balanced Robertsonian translocation. The individual carrying this unusual chromosome is unlikely to have any symptoms, however they have an increased risk of having a child with Down's syndrome. Correct identification of such translocations is therefore important for accurate counselling of recurrence risk, and requires careful selection of the appropriate genetic test (Table1). Aneuploidies can be detected with many types of genetic testing. A quantitative PCR is the fastest, and can provide a rapid diagnosis in a few days, but cannot exclude mosaic aneuploidy. Quantitative PCR will give information regarding the dosage of chromosomes, but will not give any information regarding how the genetic material is structured. Karyotype analysis – where the physical structure of chromosomes is interrogated – is useful for identifying any underlying translocations, and is therefore particularly important in cases of recurrent aneuploidy (Table 1).

2.1.2 Uniparental disomies

When a child has a uniparental disomy, their cells contain the correct number of chromosomes, however both copies of a chromosome have been inherited from the same parent. This usually occurs as a consequence of rescue of lethal trisomies caused by non-disjunction events. For most genes, this does not affect their expression. However, approximately 100 genes are subject to genomic imprinting, where their expression is strictly regulated in a parent-of-origin manner. For example, some genes are expressed only from the maternally-inherited allele, and others from the paternally-inherited allele. In uniparental disomy, the inheritance of two chromosomes from the same parent will alter the expression dosage of imprinted genes on that chromosome, which can perturb development. For example, trisomy of chromosome 15 is lethal, however the correction of such a trisomy in the early embryo resulting in paternal uniparental disomy of chromosome

15 causes Angelman Syndrome, while maternal uniparental disomy of chromosome 15 causes Prader Willi Syndrome. This can be detected using NGS approaches or by specifically looking for methylation marks which differ at imprinted regions based on which parent they were inherited from using methylation sensitive MLPA (multiplex ligation dependent PCR amplification).

2.1.3 Copy number variants

Copy number variants (CNVs) refer to deletions or duplications of parts of chromosomes, and are estimated to account for about 14% of developmental disorders. However, not all CNVs cause disease, and many CNVs show variable penetrance – not all individuals carrying the CNV will have the disease – and variable expressivity – there is variability in the severity of the associated phenotype. There are several factors that affect whether a CNV is likely to be pathogenic. In general, large and gene-rich CNVs are more likely to be pathogenic, and deletions are often more damaging than duplications. Currently, CNVs are tested for clinically by microarray, targeted FiSH, MLPA, and, if large, karyotype analysis (Table 1). Recent improvements in analysis techniques applied to exome and whole genome sequencing data are proving useful in the detection of CNVs, particularly if they are small, however these are not yet in routine clinical use.

The trinucleotide repeat expansion disorders are caused by the increased copy-number of a trinucleotide repeat, and deserve particular mention due to the specific clinical tests required to diagnose them. They are an important cause of neurodevelopmental and neuromuscular disorders, and include Fragile X syndrome - caused by the expansion of a $(CGG)_n$ repeat, Friedrich's Ataxia – $(GAA)_n$, and Myotonic Dystrophy type 1 – $(CTG)_n$. It is important to appreciate that these disorders are not yet easily identified by sequencing technologies including NGS, and instead a specific triplet primed PCR assay is used (Table 1).

2.2 Monogenic disorders

A rapidly expanding number of genes have been reported to be associated with developmental disorders. SNVs, small insertions or deletions, or CNVs that alter the function of these genes interfere with normal developmental processes. Historically, recurring phenotypic patterns were grouped into syndromes by clinical assessment, and if the

causative gene had been identified, a genetic diagnosis could be sought by Sanger sequencing. However, some clinical presentations such as early infantile epileptic encephalopathy or intellectual disability can be caused by a very large number of genes, with few distinguishing clinical features. As the cost of NGS technologies has dropped, it is now cheaper to sequence an exome, the protein-encoding portion of the genome, than to sequence more than a few genes using Sanger sequencing. Analysis of the exome data is then often limited to a list of genes on a virtual gene panel known to be associated with a given clinical presentation (e.g. neuronal migration disorders). Consequently, NGS based approaches – encompassing exomes, exome-based gene-panels and whole-genome sequencing – are increasingly being used in clinical practice. Gene agnostic approaches such as exome or whole-genome sequencing have resulted in a widening of the phenotypic spectrum of many clinically defined genetic syndromes, and have demonstrated that multiple genetic diagnoses may be found in up to 5% of patients. However, these techniques dramatically increase the amount of data generated, making analysis substantially more complex, increasing the risk of false positives, false negatives and VUS (see Table 1 and Box 1).

2.2.1 Challenges in interpreting genetic variants

As we all carry 4-5 million genetic variants (sites where our genome differs from the reference sequence), it can be challenging to identify whether an observed genetic change is responsible for a phenotype, or whether it is part of normal inter-individual variation. Widespread adoption of the American College of Medical Genetics and Genomics (ACMG) guidelines on variant interpretation has fostered a systematic approach to variant classification by clinical scientists (Table 2). Candidate diagnostic findings are best discussed with a multidisciplinary team, including a clinical geneticist or clinician with clinical and genomic expertise in the disorders being assessed. Sequence variants which are predicted to severely disrupt a protein-coding sequence, for example by the introduction of a stop codon or a frame-shift are called loss-of-function (LoF) variants and are more likely to cause disease than variants which do not alter the amino acid sequence - synonymous variants. However, LoF variants in many genes do not appear to cause disease. Missense variants change the amino acid sequence, and predicting whether this alters the function of the protein can be difficult. There are many computational algorithms which aim to predict the

consequence of missense variants, and these are utilised in clinical practice, however their results can be contradictory. The use of population genetic databases such as the Genome Aggregation Database (GnomAD) to prioritise variants has therefore become essential. Variants which are common among healthy adult members of the general population are highly unlikely to be responsible for a monoallelic rare, severe developmental disorder. Family based analysis – where both the child and their parents are sequenced in a trio – is also invaluable. If the parents are not affected by a severe developmental disorder, it is likely that the genetic cause will be either a new variant that has arisen in the child – a *de novo* variant in a dominant or haploinsufficient gene – or the inheritance of two copies of a damaging variant in a recessive gene but other possibilities such as X-linked inheritance and inherited variants in imprinted genes need to be borne in mind. A good knowledge of genetics is important in interpreting results from genetic tests and knowing whether possible causes may have been missed either by the choice of test or its analysis. Despite these steps, VUS are a common outcome of genetic tests, and can cause significant anxiety for patients. Many variants classified as 'likely pathogenic' will have a 10% risk of not being disease-causing so further assessment is usually indicated especially if any clinical or reproductive decisions are likely to be based on the diagnosis (Table 2). Some individuals with pathogenic variants will be asymptomatic if the disorder shows incomplete penetrance or may have only very subtle signs if there is highly variable expressivity. Furthermore, it is unclear whether our duty of care extends to the regular re-analysis of VUS in the light of new information, shown in research studies to yield clinical benefit, and if so, how this is to be accommodated within health care services.

2.2.2 Developmental disorders and *de novo* genetic variants

Studies such as Deciphering Developmental Disorders (DDD) have demonstrated that the genomes of children with severe developmental disorders are enriched for damaging genetic variants that have occurred *de novo*, i.e. they have arisen for the first time in the child, and have not been inherited from either parent. DDD carried out exome sequencing on ~14,000 children with severe developmental disorders and their parents. This large sample size enabled an estimation of the prevalence of severe developmental disorders caused by damaging *de novo* variants as between 1 in 213 and 1 in 448, depending on

parental age. The risks are lowest for younger parents and highest for older parents with the age of the father having a larger effect than that of the mother.

2.2.3 Recessive developmental disorders

The likelihood of a developmental disorder being recessive increases if there are affected siblings and if an individual has large stretches of autozygosity – segments of DNA that are identical (homozygous by descent) as a result of inheritance. A detailed family history, and careful enquiry as to whether there is any consanguinity in the family is important. The proportion of developmental disorders caused by recessive genetic variants is not known, however it was estimated in the DDD study to be approximately 4% among individuals of European ancestry, and 30% among individuals of Pakistani ancestry, in whom a higher burden of autozygosity was identified. However, 6% of individuals with autozygosity indicative of a first-cousin union or closer were found to have a plausible causative *de novo* variant, which underlines the importance of looking for damaging *de novo* variation even among individuals with a history of consanguinity.

3. Diagnosing the undiagnosed

Cohort studies typically find that whole exome or whole genome sequencing identifies a diagnosis for ~40% of children with severe developmental disorders. Greater severity of developmental delay, multi-system involvement and neuromuscular features tend to be associated with a higher rate of diagnosis. There are considerable international research efforts ongoing trying to identify what are the underlying causes for the remaining ~60% of children. A minority are likely to be caused by rare genetic variation in currently unknown developmental disorder genes. However, it is likely that we will need to look beyond monogenic patterns of inheritance if we are to uncover all the aetiologies of developmental disorders. Other possible causes include oligogenic and polygenic inheritance, non-coding variants and environmental factors.

3.1 Identifying new developmental disorder genes

Traditionally, new developmental disorder genes have been identified by gathering groups of similarly affected individuals and searching for a commonly affected gene – a phenotype driven approach. However, with the falling cost of NGS, it has been possible to identify new

developmental disorders through a purely statistical genotype-driven approach. As individual developmental disorders are extremely rare, very large cohort sizes are necessary to have sufficient statistical power. However, this technique has been demonstrated to work when analysing disparate developmental disorders together. In simplistic terms, the number and type of genetic variants expected in each gene can be estimated, using large databases of the genetic variation observed in the normal, healthy population – such as GnomAD. In a cohort of individuals with developmental disorders, the genes responsible will have more damaging genetic variants (variants that result in a loss-of-function of the protein, and missense variants) than expected. This allows new genetic disorders to be identified in an unbiased manner. Given the rarity of severe developmental disorders, international sharing of data between large patient cohorts will be required in order to maximise the clinical benefits, improve our understanding of the genetic architecture of developmental disorders and achieve diagnoses for as many patients as possible. These considerable benefits must be balanced against our duty of confidentiality to patients and the risks of highly identifiable personal genomic information being made public (see Box1).

3.2 Polygenic inheritance

In polygenic inheritance many different genetic loci contribute to the expression of a trait. Examples include height, weight and BMI. The majority of the genetic variants involved are common, in the non-coding portion of the genome, and each makes a small impact on phenotypic expression, which is modified by environmental factors. Until recently, it was thought that this sort of genetic variation was unlikely to play a significant role in rare, severe developmental disorders, thought to be monogenic. However, a recent study by Niemi et al. has suggested that approximately 8% of the variance in risk for neurodevelopmental disorders can be attributed to common genetic variants. This is similar to that which has been reported for common disorders such as depression. In this study there was no difference in the contribution of common genetic variation between individuals in whom a monogenic cause had been identified and those in whom it hadn’t. This suggests that common genetic variation contributes to the overall risk for developmental disorders, and may also modify the expression of individual phenotypes.

3.3 Oligogenic inheritance

This describes a situation where the inheritance of variants in several genes is required to produce a phenotype. For example, fascioscapulohumeral dystrophy type 2 requires both a rare SNV in the gene SMCHD1, and a permissive DUX4 allele for expression of the disease phenotype. Digenic models have also been identified in Bardet-Biedl syndrome, retinitis pigmentosa and midline craniosynostosis. Furthermore, in hereditary neuropathies, an increased burden of rare non-synonymous variation in neuropathy associated genes may modify the expression of other highly penetrant alleles, suggesting that mutational load may contribute to disease. This pattern of genetic aetiology can be challenging to identify but may underlie many of the currently undiagnosed cases.

3.4 Non-coding genetic variants

Non-coding DNA refers to all genetic sequence which does not encode for proteins. This is the vast majority of our DNA. Previously termed "junk" DNA, much of the non-coding genome is now thought to regulate expression of the coding sequence, controlling where, when and how genes are turned on and off. Establishing the functional significance of non-coding genetic variation is challenging for two major reasons. Firstly, our knowledge of how regulatory elements work remains largely rudimentary. It is therefore often difficult to predict which variants might disrupt gene expression, unlike in coding regions, when we are able to predict the consequence of a variant on the encoded amino acids. Secondly, it is not straightforward to identify the target gene(s) of non-coding regions, so it is difficult to assign the variant to the gene that it affects. However, the disruption of regulatory regions by CNVs resulting in perturbed gene expression and developmental disorders have been reported. Furthermore, SNVs in the *Shh* regulatory region on chromosome 7q36 have also been shown to result in abnormalities of limb bud development, including pre-axial polydactyly, tri-phalangeal thumb and syndactyly. However, a more general understanding of the contribution of non-coding variation to rare developmental disorders has been elusive. A recent study by Short *et al.* attempted to look at the contribution of *de novo* genetic variation in non-coding regions to developmental disorders. They observed an enrichment of *de novo* mutations in highly conserved fetal brain-active enhancers in neurodevelopmental disorders. However, they estimate that only 1-3% of undiagnosed developmental disorder patients are likely to carry a pathogenic *de novo* non-coding variant in a fetal brain-active enhancer.

4. Future opportunities and challenges

NGS technologies and their application to large cohorts of children with disparate developmental disorders have revolutionised our ability to provide families with a genetic diagnosis. As discussed above, there is still progress to be made in identifying the aetiologies responsible for those children currently without a genetic diagnosis. Another significant challenge is in variant interpretation. As large-scale high-resolution genetic tests are increasingly used in clinical practice, so the number of VUS is accelerating. VUS often cause significant anxiety for patients and their families, and it is unclear whether our clinical duty of care extends to regular re-analysis of previously reported VUS in the light of new data, and if so, how this can be accommodated within health care services. New technologies that allow the empirical high-throughput testing of the functional consequences of large numbers of variants in cellular systems may substantially improve our ability to interpret variants, with significant clinical benefits. Finally, genomic technologies are opening up a new era of possibilities in the treatment of rare disease. There are inspiring examples where new genetic diagnoses have suggested a mechanism and targeted treatment, resulting in startling clinical improvements for patients. Using genomic technologies to stratify clinical trial patient cohorts may reduce heterogeneity in treatment response, and so allow the development of effective targeted treatments for specific aetiologies. Although substantial technical, safety and ethical issues remain, in the longer term, gene editing and gene therapy approaches hold the promise of corrective therapy for certain conditions.

Conclusion

The application of genomic technologies has revolutionised how we approach the diagnosis of children with developmental disorders. The aspiration to provide a specific genetic diagnosis for each child with a monogenic condition, although not yet achieved, now seems realistic. A new era of potential therapeutic developments is just beginning, and there is good cause for the optimism. Children with developmental disorders make up a significant part of our paediatric population, as paediatricians we must work together with our clinical genetics colleagues to ensure that the full benefits of genomic technologies are realised for our patients.

BOX 1 Ethical issues of genomic testing

Genetic testing in children has always raised ethical, social and legal issues, such as informed consent in minors, best-interests decision making, reproductive autonomy, incidental findings, predictive testing and the implications of genetic data for relatives. Many of these issues are amplified by the scale and depth of genomic data, but there are also novel considerations raised by genomic scale analyses:

Balancing confidentiality and the benefits of data sharing

An individual's genome sequence is unique to them, and is therefore highly identifying. Patients are often concerned about the possibility of genetic discrimination. Anonymisation is not a sufficient safeguard, as several studies have shown that re-identification is possible using just a tiny proportion of an individual's genome. This can have significant unintended consequences for other family members, in addition to the child tested, some of which may not be foreseeable. However, given our incomplete knowledge of the genome, in order to interpret genomic data and gain maximum clinical benefit, we need to compare an individual's data against that of many others, necessitating the sharing of this data. A proportionate approach to data sharing is developing within the field. This allows for these clinical benefits to be realised through the carefully managed secure sharing of anonymised individual genomic data and detailed phenotypic descriptions with other clinicians and authorised researchers, while much more limited and less identifiable data – likely diagnostic variants with a limited phenotypic description – is shared openly through platforms such as the DECIPHER database and Matchmaker Exchange (Table 3), facilitating the identification of new genetic conditions, and realising considerable patient benefit.

Incidental findings

Incidental findings are commonly encountered in many diagnostic tests, and should be routinely discussed when obtaining consent. However, genetic data has the potential to give predictive information – for example, regarding a patient's future risk of developing breast cancer or dementia – and the possible consequences of this need to be considered, particularly in children who lack capacity to consent. The ACMG has identified 59 medically actionable genes in which they recommend feeding back damaging genetic variants to the adult patient as secondary findings when sequencing has been undertaken for other

reasons. However, there is no prospective data on the benefits and risks of such opportunistic screening. As current clinical practice is to not undertake predictive testing in a child for any later-onset condition until they have the capacity to consent, only information regarding childhood-onset conditions are returned. However, it is unclear whether our clinical duty of care extends to reassessing this when the child reaches the age of majority, and if so, how to deliver this within the health care system.

BOX2 Taking consent for genetic tests

Often the child does not have capacity to consent for themselves, so consent is given by the parents in the best interest of the child. However, where possible the child should be involved in this process. The broad domains that should be covered to ensure that consent is informed and valid are below. Consent should be taken by the most appropriate clinician, and for many genetic tests this will remain a clinical geneticist. However, consent for microarray testing is increasingly being taken by paediatricians. Therefore, we will discuss the principles of consent in the context of the following clinical vignette: a baby is born with a cleft palate and tetralogy of Fallot.

1. What the test checks for and what it cannot exclude:

"A microarray checks for big pieces of missing or duplicated DNA in our "instruction manual". We know that sometimes this can cause heart and palate problems. However, if the test is normal, there might still be a genetic cause. This is because the test does not check for certain genetic changes – there may still be small sentences missing, or spelling mistakes in the genetic information. We may therefore suggest doing further tests in the future. "

2. Variants of uncertain significance:

"Our understanding of the human genetic code is not complete, and therefore sometimes we find things where we are unsure what they mean. For some genetic changes, we may not be able to give you a clear answer as to whether it's the cause of your son's heart defect and cleft palate. We call these variants of uncertain significance. We may need to take blood tests from yourselves, the parents, to try to interpret the results."

3. Incidental findings

“We are doing this test to see if we can find a reason why your baby has a cleft palate and a problem with his heart. However, by doing this test, we sometimes find additional information that we weren’t looking for – for example, we may discover that your baby may have an increased risk of learning difficulties.”

Table 1 – Common Genetic tests, their resolution and clinical application.

Test	Resolution	No. of loci tested	Variants detected	Variants expected per person	Clinical advantages	Clinical disadvantages	Example clinical question
Quantitative PCR for chromosome count	Chromosome presence / absence	N/A	Aneuploidy	0-1	Fast – result usually within 3 days	Only available for specific chromosomes (13, 18, 21, X, Y)	Does this baby have Trisomy 21?
G-banded Karyotype	5-10Mb	~500	Aneuploidy, polyploidy, CNVs > 5Mb	0-1	Looks at structure – translocations that confer risk of recurrent aneuploidy will be identified	Small CNVs will not be seen.	A baby with Trisomy 21 is born to a family history of Trisomy 21, does one parent carry a balanced translocation?
Microarray / Comparative genomic hybridisation (CGH) array.	5-100Mb	2 million	CNVs	10-100s	Smaller CNVs identified.	Genetic dosage but not structure is interrogated. If a deletion/duplication identified at the end of a chromosome, advisable to perform a karyotype to look for translocations.	A baby is born with a cleft palate and tetralogy of Fallot. Does this baby have 22q11.2 deletion syndrome?
Fluorescent *in situ* hybridisation (FiSH)	10kb – whole chromosome	N/A	Aneuploidy and CNVs	0-1	Fast – result usually within 1 week	A specific test must be developed and validated for each region, therefore it is only available for some conditions.	A baby is born with a cleft palate and tetralogy of Fallot. Does this baby have 22q11.2 deletion syndrome?

MLPA	Exons within a particular gene.	Test dependent	Exon deletions Methylation sensitive MLPA can detect uniparental disomy	0-1	Highly sensitive, often paired with single gene sequencing.	A specific test must be developed and validated for each region, therefore it is only available for certain genes.	A sweat test suggests a baby has CF, but genotyping of common variants is normal. MLPA for exonic deletions + sequencing of *CFTR* is undertaken.
Triplet primed PCR assay	Specific trinucleotide repeat regions	Test dependent	Trinucleotide repeat expansions	Test dependent	Quantifies the copy-number of specific trinucleotide repeat tracts.	A specific test must be developed and validated for each region.	A 3yr old boy presents with speech delay, hyperactivity and autistic features. Triplet primed PCR is requested for Fragile X syndrome.
Sequencing based genetic tests:							
Single gene sequencing	1bp	Test dependent	Single nucleotide variants in a single gene	Test dependent	Targeted testing at base-pair resolution, minimises VUSs	1. Must have a strong clinical suspicion of what gene to test. 2. Needs to be paired with MLPA to detect exonic deletions.	A sweat test suggests a baby has CF, but genotyping of common variants is normal. MLPA for exonic deletions + sequencing of *CFTR* is undertaken.
Gene panel sequencing	1bp	10-100,000	Single nucleotide variants in	100-1000s	Very useful where the clinical phenotype can	1. Only the genes on the panel will be analysed. As more genes are associated with	A 10w baby presents with epileptic encephalopathy.

			particular genes		be caused by variants in many different genes e.g. intellectual disability	paediatric disease, panels will need to be regularly updated. 2. VUS are common.	She is developmentally delayed but not dysmorphic. An early infantile epileptic encephalopathy gene panel is sent.
Exome	1bp	~ 50 million	Coding SNVs and CNVs	~20,000	Unbiased, all genes will be tested. Analysing both parents concurrently significantly reduces VUS	1. Only ~30% of genes have a known disease association 2. Predicting the consequence of variants is not straightforward and requires both clinical and molecular assessment. 3. VUS are common 4. Trio-based approaches have the potential to reveal non-paternity. 5. Trinucleotide repeat expansions are not currently detected.	A 10w old baby presents with epileptic encephalopathy. Gene panel testing is negative, so DNA from the child and both parents is sent for trio exome sequencing
Whole genome sequencing	1bp	~3 billion	Most variants	4-5 million	Unbiased, all genes will be tested. Analysing both parents concurrently	1. Only ~30% of genes have a known disease association 2. Predicting the consequence of variants is not	A 10w old baby presents with epileptic encephalopathy. Gene panel testing is negative, so DNA

					significantly reduces VUS. CNV detection is easier than with exomes.	straightforward and requires both clinical and molecular assessment. 3. VUS are common 4. Trio-based approaches have the potential to reveal non-paternity. 5. Trinucleotide repeat expansions are not currently detected. 6. The majority of the noncoding genome is not currently interpretable.	from the child and both parents is sent for trio whole genome sequencing

Table 2: Variant classification according to the American College of Medical Genetics and Genomics guidelines

Variant class definition	**Evidence needed for variant classification (simplified)**	**How can this variant be used in the clinic?**
5 – Pathogenic or clearly pathogenic Probability of being pathogenic, $p>0.99$	- A well-established disease-causing variant with consensus on pathogenicity - Co-segregation with the disease observed in several families - Functional studies have detailed the pathogenic effect - Variant not seen or very rarely seen in healthy populations.	Together with other laboratory data, this variant can sometimes be used to predict the phenotype, and therefore can be used to direct clinical management. This includes: - Screening all relatives who may have inherited the variant - Pre-natal diagnosis However, beware incomplete penetrance and variable expressivity.
4 – Likely pathogenic Probability of being pathogenic, $0.9 < p < 0.99$	Lacks some evidence for a class 5 variant. Many disease-causing but novel variants, particularly in less studied genes will be class 4.	Could be cautiously used for clinical management, but ongoing review of variant status is required. - Could additional investigations (e.g. biochemical studies) provide new evidence? - Should not be used for predictive testing or prenatal diagnosis without discussion with a Clinical Genetics unit.
3 – Variant of uncertain significance Probability of being pathogenic $0.1 < p < 0.9$	Variant has conflicting evidence – some disease-causing characteristics, other data suggests it may be benign (e.g. seen at high frequency in a healthy population). OR Insufficient evidence to categorise the variant.	Should not be used to guide clinical decision making. - Could co-segregation (testing relatives) or functional studies facilitate reclassification to class 2 or 4? If undertaken, be clear with relatives that this is not screening for disease. - Be aware that variant classification may change over time, and should be reviewed.
2 – Likely benign Probability of being pathogenic, $0.001<p<0.1$	Variant found more commonly in the healthy population than expected for the frequency of the disease.	Normally not reported, as not clinically actionable

1 – Benign Probability of being pathogenic, p<0.001	- Variant does not segregate with disease in families with >2 affected individuals. - Functional studies demonstrate no significant effect of variant.	Normally not reported, as not clinically actionable

Table 3: Online resources and their clinical utility

Online resource	Description of resource	Example clinical questions
OMIM https://www.omim.org/ Genetics Home Reference https://ghr.nlm.nih.gov/resources	Describes what is known about clinical syndromes and human genes	- Is the causative gene known for this condition? - What phenotypes are associated with this condition? - Are any clinical disorders associated with this gene?
DECIPHER https://decipher.sanger.ac.uk ClinVar https://www.ncbi.nlm.nih.gov/clinvar HGMD http://www.hgmd.cf.ac.uk/ac/index.php	Databases of phenotypic and genotypic information.	- Are there other patients with the same or similar variants in this gene? What are their phenotypes? - Are there patients who have CNVs that remove the gene that my variant is in? - How common are damaging variants in this gene in healthy people, e.g. in GnomAD? (DECIPHER only)
Matchmaker Exchange https://www.matchmakerexchange.org	Facilitates the sharing of phenotypic and genotypic information between clinicians in an effort to 'solve' unresolved cases.	- Are there other patients with similar variants and similar phenotypes? Could this represent a new condition?

Genetic Alliance UK https://www.geneticalliance.org.uk	A coalition of organisations that support individuals and families with rare genetic conditions.	- Is there a patient support group or charity that I could direct my patient and their family to, if they are interested? - Educational resources on genetics and genetic disorders.

Take home messages:

- NGS technologies have the power to improve diagnosis and treatment of rare paediatric disease.
- Interpreting genomic data is complex.
- Making a diagnosis of a severe, life-long genetic disease in a young person is a major step and due diligence is required to ensure that the diagnosis is clinically and molecularly robust.
- Consider referral to a Clinical Geneticist or colleague with expert knowledge of the specific genetic diagnosis under consideration. Clinical experience of the presentation and progression of disease remains crucial to recognise when an erroneous diagnosis has been made or when the genetic test explains only part of the clinical presentation and there is a second important diagnosis to make.
- Approximately 5% of patients will have >1 genetic diagnosis to explain their clinical features.

Acknowledgements

We thank Drs S Sabanathan and T Rossor for helpful comments on the text.

Reading List:

1 Wright, C. F., FitzPatrick, D. R. & Firth, H. V. Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet* **19**, 325, doi:10.1038/nrg.2018.12 (2018).

2 McCandless, S. E., Brunger, J. W. & Cassidy, S. B. The burden of genetic disease on inpatient care in a children's hospital. *Am J Hum Genet* **74**, 121-127, doi:10.1086/381053 (2004).

3 Deciphering Developmental Disorders, S. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433-438, doi:10.1038/nature21062 (2017).

4 Martin, H. C. *et al.* Quantifying the contribution of recessive coding variation to developmental disorders. *Science* **362**, 1161-1164, doi:10.1126/science.aar6731 (2018).
5 Niemi, M. E. K. *et al.* Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* **562**, 268-271, doi:10.1038/s41586-018-0566-4 (2018).
6 Wright, C. F., Hurles, M. E. & Firth, H. V. Principle of proportionality in genomic data sharing. *Nat Rev Genet* **17**, 1-2, doi:10.1038/nrg.2015.5 (2016).
7 Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* **84**, 524-533, doi:10.1016/j.ajhg.2009.03.010 (2009).
8 de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367**, 1921-1929, doi:10.1056/NEJMoa1206524 (2012).
9 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
10 Short, P. J. *et al.* De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* **555**, 611-616, doi:10.1038/nature25983 (2018).
11 Lupski, J. R. Digenic inheritance and Mendelian disease. *Nat Genet* **44**, 1291-1292, doi:10.1038/ng.2479 (2012).
12 Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum Mol Genet* **24**, R102-110, doi:10.1093/hmg/ddv259 (2015).
13 Posey, J. E. *et al.* Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. *N Engl J Med* **376**, 21-31, doi:10.1056/NEJMoa1516767 (2017).
14 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).